Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 29 June 2001
Accepted 7 September 2001

A new interpretation and practical aspects of the direct-methods modulus sum function. VIII

Jordi Rius,[a]* Xavier Torrelles,[a] Carles Miravitlles,[a] J. M. Amigó[b] and M. M. Reventós[b]

[a]Institut de Ciència de Materials de Barcelona (CSIC), Campus de la UAB, 08193 Cerdanyola, Catalunya, Spain, and [b]Departament de Geologia, Universitat de València, 46100 Burjassot, Spain. Correspondence e-mail: jordi.rius@icmab.es

Since the first publication of the direct-methods modulus sum function [Rius (1993). *Acta Cryst.* A**49**, 406–409], the application of this function to a variety of situations has been shown in a series of seven subsequent papers. In this way, much experience about this function and its practical use has been gained. It is thought by the authors that it is now the right moment to publish a more complete study of this function which also considers most of this practical knowledge. The first part of the study relates, thanks to a new interpretation, this function to other existing phase-refinement functions, while the second shows, with the help of test calculations on a selection of crystal structures, the behaviour of the function for two different control parameters. In this study, the principal interest is focused on the function itself and not on the optimization procedure which is based on a conventional sequential tangent formula refinement. The results obtained are quite satisfactory and seem to indicate that, when combined with more sophisticated optimization algorithms, the application field of this function could be extended to larger structures than those used for the test calculations.

© 2002 International Union of Crystallography
Printed in Great Britain – all rights reserved

1. Introduction

Direct methods solve crystal structures by combining the information contained in the measured intensities with the positivity and atomicity of the electron-density distribution. Conventional multisolution direct methods refine the collectivity Φ of phases $\varphi_{\mathbf{h}}$ of the normalized structure factors $E_{\mathbf{h}}$ of the strong reflections $\mathbf{h}$ by searching for an extremum of a given target or phase refinement function. For simplicity, an equal-atom structure with N atoms in the unit cell belonging to space group $P1$ will be assumed throughout this paper.

From a practical point of view, there exist two very important types of phase-refinement functions:

(i) Functions minimizing the discrepancies between the cosines of the three-phase structure invariants $\Phi_3(\mathbf{h}, \mathbf{h}')$ and their expectation values ε as, for example, in the function Ψ_D of Debaerdemaeker & Woolfson (1983),

$$\Psi_D(\Phi) = \sum_{\mathbf{h}} \sum_{\mathbf{h}'} K(\mathbf{h}, \mathbf{h}')\{\cos[\Phi_3(\mathbf{h}, \mathbf{h}')] - \varepsilon[K(\mathbf{h}, \mathbf{h}')]\}^2 = \text{minimum}, \quad (1)$$

wherein the indices $\mathbf{h}$, $\mathbf{h}'$ and $\mathbf{h} - \mathbf{h}'$ denote reflections with large $|E|$ values and

$$K(\mathbf{h}, \mathbf{h}') = 2N^{-1/2}|E_{-\mathbf{h}}||E_{\mathbf{h}'}||E_{\mathbf{h}-\mathbf{h}'}| \quad (2)$$

is the concentration parameter of the conditional probability distribution of the triplets. This function was tested by Debaerdemaeker & Woolfson (1983) with rather poor results. As shown later by Weeks *et al.* (1993), the effectiveness of this type of function increases dramatically when the phase refinement in reciprocal space alternates with real-space calculations to impose atomic constraints. This iterative two-step refinement is known as the 'Shake-and-Bake procedure' and is responsible for the fully automatic routine solution of structures in the 200–400 atom range (Weeks *et al.*, 1994). The phase refinement is normally carried out by using either a modified tangent formula or the parameter-shift method (Bhuiya & Stanley, 1963) adapted to phase refinement.

(ii) A second type of phase-refinement function is based on the minimization of the residual

$$R(\Phi) = \langle[|E_{\mathbf{H}}| - |E_{\mathbf{H}}(\Phi)|]^2\rangle_{\mathbf{H}} \quad (3)$$

with $\mathbf{H}$ denoting both large and weak reflections, and where the atomicity constraint is introduced *via* Sayre's equation by making in expression

$$|E_{\mathbf{H}}(\Phi)| = c_{\mathbf{H}} \sum_{\mathbf{h}'} |E_{\mathbf{h}'}||E_{\mathbf{H}-\mathbf{h}'}| \cos(\phi_{-\mathbf{H}} + \varphi_{\mathbf{h}'} + \varphi_{\mathbf{H}-\mathbf{h}'}) \quad (4)$$

the respective phases φ and ϕ of the structure factors of the strong reflections in the true and squared structures equal (Rius, 1993).

The residual can be written in the form

$$\begin{aligned} R(\Phi) &= A - 2B(\Phi) + C(\Phi) \\ A &= \langle |E_{\mathbf{H}}|^2 \rangle_{\mathbf{H}} \\ B(\Phi) &= \langle |E_{\mathbf{H}}||E_{\mathbf{H}}(\Phi)| \rangle_{\mathbf{H}} \\ C(\Phi) &= \langle |E_{\mathbf{H}}(\Phi)|^2 \rangle_{\mathbf{H}} \end{aligned} \tag{5}$$

which clearly shows that minimization of $R(\Phi)$ is equivalent to minimizing the difference $C(\Phi) - 2B(\Phi)$. By neglecting the contributions of the weak reflection to $B(\Phi)$ and since $C(\Phi)$ can be expressed in the form

$$C(\Phi) = \sum_{\mathbf{h}} E_{-\mathbf{h}} \left\langle c_{\mathbf{H}}^2 E_{-\mathbf{H}+\mathbf{h}} \sum_{\mathbf{h}''} E_{\mathbf{h}''} E_{\mathbf{H}-\mathbf{h}''} \right\rangle_{\mathbf{H}} \tag{6}$$

Rius & Miravitlles (1991) showed that $R(\Phi)$ can be minimized by means of the modified tangent formula

$$\varphi_{\mathbf{h}} = \text{phase of } \left\{ \sum_{\mathbf{h}'} E_{\mathbf{h}'} E_{\mathbf{h}-\mathbf{h}'} - c \sum_{\mathbf{H}} E_{-\mathbf{H}+\mathbf{h}} \sum_{\mathbf{h}''} E_{\mathbf{h}''} E_{\mathbf{H}-\mathbf{h}''} \right\} \tag{7}$$

which is closely related to the Sayre-equation tangent formula (Debaerdemaeker *et al.*, 1985). However, one difficulty for its practical application comes from the presence of the scaling constant c, the average value of $c_{\mathbf{H}}$. At this point, the question arises whether there exists an alternative way of expressing R that leads to a phase-refinement function without the scaling constant problem. As will be shown in the next section, the direct-methods modulus sum function offers one solution.

2. A new interpretation of the modulus sum function $S(\Phi)$

The residual $R(\Phi)$ can be expressed in an alternative way by making full use of the atomicity constraint. Effectively, if the atomicity condition is fulfilled, then, as predicted by the acentric probability distribution of the $|E|$ values, the moduli $|E|$ have restricted values close to 1 (Wilson, 1949; Woolfson, 1970). Consequently, the following approximate relationship between the moduli $|E|$ and their squares $|E|^2$ will hold (Fig. 1), namely

$$\begin{aligned} |E|^2 &\cong \langle |E|^2 \rangle + (|E| - \langle |E| \rangle)(\mathrm{d}|E|^2/\mathrm{d}|E|)_{|E|=\langle |E| \rangle} \\ &= 2\langle |E| \rangle |E| - 2\langle |E| \rangle^2 + \langle |E|^2 \rangle. \end{aligned} \tag{8}$$

By introducing in $|E|$ the dependence on Φ and taking the average over $\mathbf{H}$, $C(\Phi)$ may be approximated by

$$C(\Phi) \cong 2\langle |E_{\mathbf{H}}| \rangle_{\mathbf{H}} \langle |E_{\mathbf{H}}(\Phi)| \rangle - 2\langle |E_{\mathbf{H}}| \rangle_{\mathbf{H}}^2 + \langle |E_{\mathbf{H}}|^2 \rangle_{\mathbf{H}}. \tag{9}$$

Next, by replacing (9) in (5), the residual expression simplifies to

$$R(\Phi) = S_{\text{exp}} - S(\Phi) = \text{minimum} \tag{10}$$

wherein

$$S_{\text{exp}} = 2(\langle |E_{\mathbf{H}}|^2 \rangle_{\mathbf{H}} - \langle |E_{\mathbf{H}}| \rangle_{\mathbf{H}}^2) \tag{11}$$

$$S(\Phi) = 2\langle [|E_{\mathbf{H}}| - \langle |E_{\mathbf{K}}| \rangle_{\mathbf{K}}] |E_{\mathbf{H}}(\Phi)| \rangle_{\mathbf{H}}. \tag{12}$$

$S(\Phi)$ is the so-called direct-methods modulus sum function (Rius, 1993; Rius, Torrelles & Miravitlles, 2000). From (10), it follows that, as long as the atomicity condition is satisfied, minimization of $R(\Phi)$ is equivalent to maximizing $S(\Phi)$ and that no estimation of scaling constants is necessary. By use of the same procedure as Debaerdemaeker *et al.* (1985), the new phase estimates maximizing $S(\Phi)$ can be solved for the limit of $S(\Phi)$,

$$\begin{aligned} (\partial/\partial\varphi_{\mathbf{h}}) S(\Phi) &= 2\langle [|E_{\mathbf{H}}| - \langle |E_{\mathbf{K}}| \rangle_{\mathbf{K}}] (\partial/\partial\varphi_{\mathbf{h}}) |E_{\mathbf{H}}(\Phi)| \rangle_{\mathbf{H}} \\ &= 0 \qquad \text{for all } \mathbf{h}, \end{aligned} \tag{13}$$

which leads to the sum function tangent formula, SF-TF (Rius, 1993),

$$\begin{aligned} \varphi_{\mathbf{h}} = \text{phase of } \Big\{ &\sum_{\mathbf{h}'} X_{\mathbf{h},\mathbf{h}'} E_{\mathbf{h}'} E_{\mathbf{h}-\mathbf{h}'} \\ &+ \sum_{\mathbf{l}} [|E_{\mathbf{l}}| - \langle |E_{\mathbf{K}}| \rangle_{\mathbf{K}}] E_{\mathbf{h}-\mathbf{l}} \exp \phi_{\mathbf{l}} \Big\} \end{aligned} \tag{14}$$

with

$$X_{\mathbf{h},\mathbf{h}'} = 1 - \tfrac{1}{3} \langle |E_{\mathbf{K}}| \rangle_{\mathbf{K}} [(1/E_{-\mathbf{h}}) + (1/E_{\mathbf{h}'}) + (1/E_{\mathbf{h}-\mathbf{h}'})]. \tag{15}$$

In (14), the sum over $\mathbf{h}'$ takes into account the contributions of the large reflections while the sum over $\mathbf{l}$ considers the contributions of the weak ones.

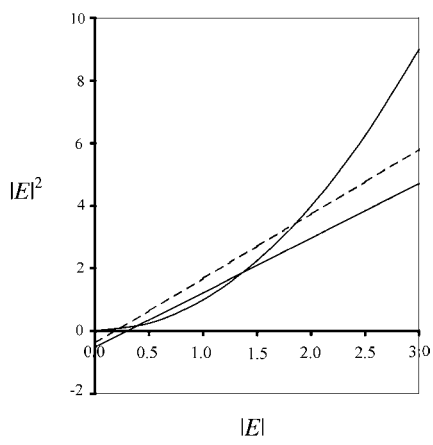

Figure 1
$|E|^2$ approximated with equation $2\langle |E| \rangle_{\mathbf{H}} |E| - 2\langle |E| \rangle_{\mathbf{H}}^2 + \langle |E|^2 \rangle_{\mathbf{H}}$ (straight lines) using experimental mean values taken from MBH2; dashed line for $N_{\text{weak}} \cong N_{\text{large}}$; continuous line with all reflections.

3. Principal advantages and limitations of function $S(\Phi)$

(i) The principal limitation of $S(\Phi)$, which also holds for $\Psi_{\mathbf{D}}(\Phi)$, is the fulfilment of the atomicity condition, *i.e.* the availability of intensity data up to atomic resolution.

(ii) One important advantage of $S(\Phi)$ is the simplicity of implementation. As has been pointed out above, it involves no scaling constants or complex weighting schemes, and in addition it does not require alternate refinements in reciprocal and real space. This function is quite useful, for example, for solving structures from powder data (Rius *et al.*, 1995, 1999; Rius, Torrelles, Miravitlles, Ochando *et al.*, 2000; Rius, 2000) and, as shown in the test examples, it is also capable of determining structures with up to 500 atoms in the asymmetric

Table 1
Relevant information for the compounds used in the test calculations.

PDB is the file code in the Protein Data Bank. The composition of the asymmetrical unit is given as the approximate number of C, N and O atoms followed by the additional non-H atoms (if any). The selected crystal structures cover a variety of space groups and a variable degree of difficulty in the application of direct methods.

Code	Space group	Composition	Z	Reference	PDB
PGE2	$P1$	25	1	DeTitta *et al.* (1980)	–
MBH2	$P1$	18	3	Poyser *et al.* (1986)	–
SUOA	$P2_12_12_1$	47	4	Oliver & Strickland (1984)	–
Goldman2	Cc	28	8	Irngartinger *et al.* (1981)	–
MGHEX	$P3_1$	92 + Mg + 2Cl	3	Williams & Lawton (1975)	–
Munich1	$C2$	20	8	Szeimies-Seebach *et al.* (1978)	–
BHAT	Pc	42	2	Bhat & Ammon (1990)	–
TVAL	$P1$	78	2	Smith *et al.* (1975)	–
Winter2	$P2_1$	82 + 2Cl	2	Butters *et al.* (1981)	–
APAPA	$P4_12_12$	67 + 2P	8	Suck *et al.* (1976)	–
PEP1	$P2_12_12_1$	85	4	Antel *et al.* (1995)	–
APP	$C2$	300 + Zn	4	Glover *et al.* (1983)	–
Rubredoxin	$P2_1$	452 + Fe	2	Sheldrick *et al.* (1993)	8rxn
Alpha1	$P1$	503 + Cl	1	Privé *et al.* (1999)	1byz
Pheromone	$C2$	318 + 7S	4	Anderson *et al.* (1996)	2erl

unit. In a slightly modified form, it is also able to cope with the in-plane diffraction intensity data of reconstructed surfaces (Rius *et al.*, 1996; Torrelles, Rius, Miravittles & Ferrer, 1998; Torrelles, Rius, Boscherini *et al.*, 1998; Pedio *et al.*, 2000).

(iii) One of the strengths of $S(\Phi)$ comes from the active use of the weak E values, which is decisive for the phase-refinement effectiveness without leaving the reciprocal space. This fact, however, turns into a drawback for certain applications, *e.g.* the solution of the anomalous-scatterer substructure in proteins. Since the reflections with weak substructure contributions are here unknown, $S(\Phi)$ cannot be applied (at least at present). The Shake-and-Bake algorithm avoids this problem by combining refinement of $\Psi_D(\Phi)$ (which only requires large $|E|$ values) with the application of the atomic constraints in real space.

(iv) $S(\Phi)$ uses triplets of type *s-s-s* and *w-s-s* (*s* = strong, *w* = weak). For small and medium sized structures, this is not a problem. For large structures, however, the number of triplets becomes very high, so that the function $\Psi_D(\Phi)$, which only involves triplets of type *s-s-s*, seems to be more favourable. Recently, Burla *et al.* (2000) have presented a new direct-methods strategy which consists of applying, for a given trial, successive cycles of phase refinement while gradually reducing the $|E|_{\min}$ value. The resulting direct-methods solution is exhaustively worked out in real space. Unlike 'Shake and Bake', direct methods do not alternate with the real-space part, so that a large number of triplets has a less negative effect in this new strategy.

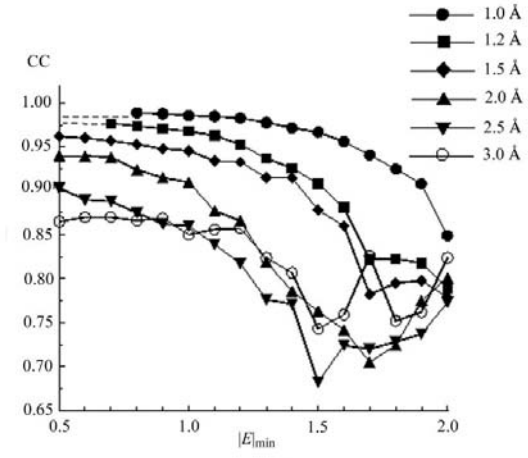

Figure 2
Correlation coefficient CC between $|E_{\mathbf{h}}|$ and $X_{\mathbf{h}} = |E_{\mathbf{h}}(\Phi)| \cos(\varphi_{-\mathbf{h}} + \phi_{\mathbf{h}})$ for different cut-off values $|E|_{\min}$ and data resolutions $d_{\min}$. $X_{\mathbf{h}}$ computed with correct phases. Test data taken from a purely organic compound. For $d_{\min} \leq 1.0$ Å, CC is relatively insensitive to the cut-off value, so that values as large as $|E|_{\min} \cong 1.6$ can be tolerated. For $d_{\min} > 1$ Å, deterioration progresses very quickly.

4. Optimizing the control parameters

Two important parameters in the implementation of function $S(\Phi)$ have been investigated with the help of some test calculations: (i) the value of $|E|_{\min}$, *i.e.* the cut-off value for considering a reflection as large, and (ii) once the optimum $|E|_{\min}$ is fixed, the ratio r of number of weak reflections to number of large reflections ($r = N_{\rm weak}/N_{\rm large}$).

4.1. The cut-off value $|E|_{\min}$

This value controls the number of large reflections in $S(\Phi)$ and, consequently, the number of phases to be refined. It also has a direct influence on the accuracy of $|E_{\mathbf{H}}(\Phi)|$ since, as indicated in (4), $|E_{\mathbf{H}}(\Phi)|$ is expressed as a function of Φ using Sayre's equation. It is tempting to select a large $|E|_{\min}$ to reduce as much as possible the number of refined phases. However, since this is at the cost of the accuracy of $|E_{\mathbf{H}}(\Phi)|$, it is first necessary to study to what extent the variation of $|E|_{\min}$ affects the accuracy at different data resolutions $d_{\min}$. This

Table 2
Results of the application of functions $S(\Phi)$ to the test structures.

$d_{\min}$ is the upper resolution limit in Å of the data used; $|E|_{\min}$ is the cut-off value for considering a reflection as large; Φ is the number of refined phases; $\langle|E|\rangle$ is the mean value of $|E|$ taken over all reflections taking part in the refinement; C-TF is the classical tangent formula (Karle & Hauptman, 1956) and SF-TF the tangent formula maximizing the modulus sum function (Rius, 1993), number of trials at the right; r.n. is the ranking number of the cluster with the correct trials and R the corresponding 'pseudo-residual' figure of merit (see text for definition) obtained after Fourier recycling. Inspection of the table clearly shows that the best results are obtained for $|E|_{\min} = 1.25$. Not only the ranking numbers of the correct solutions are lower but also the number of correct solutions increase. In most cases, the correct solution is the top solution.

Code	$d_{\min}$	$\|E\|_{\min}$	Φ	$\langle\|E\|\rangle$	% sol. with C-TF	% sol.with SF-TF	r.n.	R (%)
PGE2	1.0	1.25	169	1.00	0 / 100	94 / 100	1	38†
		1.45	112	1.09	0 / 100	39 / 100	1	
		1.65	78	1.17	0 / 100	64 / 100	1	
MBH2	1.0	1.25	448	0.97	0 / 100	88 / 100	1	15
		1.45	278	1.03	0 / 100	69 / 100	1	
		1.65	162	1.10	0 / 100	2 / 100	12	
SUOA	1.0	1.25	385	0.96	0 / 100	5.2 / 100	4	13
		1.45	240	1.03	0 / 100	2 / 100	1	
		1.65	152	1.12	0 / 100	0 / 100	–	
Goldman2	1.0	1.25	310	0.95	0 / 100	92 / 100	1	11
		1.45	197	1.02	0 / 100	49 / 100	1	
		1.65	112	1.10	0 / 100	0 / 100	–	
MGHEX	1.0	1.25	677	0.93	0 / 100	29 / 100	5	23
		1.45	374	0.99	0 / 100	24 / 100	2	
		1.65	196	1.06	0 / 100	0 / 100	–	
Munich1	1.0	1.25	313	0.96	0 / 100	4 / 100	1	20
		1.45	206	1.02	0 / 100	5 / 100	8	
		1.55	163	1.06	1 / 100	0 / 100	–	
BHAT	1.0	1.25	221	0.98	0 / 100	44 / 100	1	15
		1.45	136	1.05	1 / 100	15 / 100	4	
		1.65	74	1.15	0 / 100	0 / 100	–	
TVAL	1.0	1.25	754	1.01	0 / 100	100 / 100	1	14
		1.45	423	1.11	2 / 100	82 / 100	1	
		1.65	245	1.21	0 / 100	7 / 100	5	
Winter2	1.0	1.25	669	0.93	0 / 100	3 / 100	1	20
		1.50	353	1.03	0 / 100	2 / 100	2	
		1.60	256	1.08	0 / 100	0 / 100	–	
APAPA	1.0	1.50	307	1.00	1 / 100	7 / 100	4	28
		1.65	204	1.06	0 / 100	1 / 100	2	
PEP1	1.0	1.45	407	1.01	0.2 / 1000	1.5 / 1000	2	21
		1.65	242	1.09	0 / 1000	0 / 1000	–	
APP	1.10	1.45	1087	0.95	–	7 / 100	1	43
Rubredoxin	1.15	1.45	1121	0.99	–	1 / 100	3	43
Alpha1	1.00	1.59	1636	1.08	–	1.4 / 500	1	25
Pheromone	1.05	1.50	1068	1.04	–	0.025 /4000	65	48

† The relatively large R value is due to the large thermal disorder in the chains.

study has been performed with the data of the organic structure PEP1. The indicator selected for measuring the evolution of the accuracy for different $|E|_{\min}$ and $d_{\min}$ values has been the correlation coefficient CC between the values $|E_{\mathbf{h}}|$ and $X_{\mathbf{h}} = |E_{\mathbf{h}}(\Phi)| \cos(\varphi_{-\mathbf{h}} + \phi_{\mathbf{h}})$, the latter calculated with correct phases (Fig. 2),

$$\mathrm{CC} = \left[\sum |E_{\mathbf{h}}| X_{\mathbf{h}}\right] \Big/ \left[\sum |E_{\mathbf{h}}|^2 \sum X_{\mathbf{h}}^2\right]^{1/2}. \qquad (16)$$

The principal conclusion from inspection of Fig. 2 is that, for $d_{\min} \cong 1$ Å, $|E|_{\min}$ values up to approximately 1.6 can be tolerated (CC = 0.96). At this resolution, the degradation of the accuracy progresses relatively slowly. This is in clear contrast with the situation for $d_{\min} \cong 1.2$ Å. Here, the accuracy deteriorates markedly as evidenced by the value CC = 0.89 obtained for $|E|_{\min} = 1.55$. This low CC value means that, at this resolution, purely organic compounds will be hardly solved with the $S(\Phi)$ function.

Once the upper cut-off value of $|E|_{\min}$ is fixed at 1.6–1.7 for $d_{\min}$ values close to 1 Å, tests with smaller $|E|_{\min}$ values on a variety of diffraction data sets were performed. Their degree of difficulty is variable and in most cases the classical tangent formula (Karle & Hauptman, 1956; Yao, 1981) cannot solve them in a reasonable number of trials. The diffraction data are almost complete up to $d_{\min} = 1$ Å and the number of weak reflections has been estimated from $N_{\mathrm{weak}} \cong N_{\mathrm{large}}$. The information about the selected compounds is summarized in Table 1 which also contains the PDB file code when protein data have been deposited in the Protein Data Bank. The largest structures are at the bottom of the table. All calculations have been carried out with a new version of program *XLENS*™ which, after clustering and sorting the direct-methods solutions according to the refined $S(\Phi)$ values, automatically performs, for each solution, a Fourier-recycling step followed by the corresponding R-value calculation for the measured reflections. The direct-methods solutions are analysed until the correct solution is found or a given cut-off

Table 3
Success rates (%) of function $S(\Phi)$ for different ratios $r = N_{\rm weak}/N_{\rm large}$ where $N_{\rm weak}$ and $N_{\rm large}$ are the numbers of weak and large reflections.

All calculation were performed with $|E|_{\rm min} = 1.25$. The principal conclusion is that the success rates of $S(\Phi)$ are rather insensitive to ratio variations within the interval studied.

Code	$N_{\rm large}$	$r = 0.5$	$r = 1.0$	$r = 1.5$	$r = 2.0$
PGE2	169	65.0	83.0	65.0	94.0
MBH2	448	81.0	98.0	99.0	96.0
SUOA	385	3.0	5.5	4.2	2.5
Goldman2	310	97.0	95.0	98.0	88.0
MGHEX	677	40.0	23.0	27.0	18.0
Munich1	313	4.4	3.8	7.6	10.20
BHAT	221	23.0	44.0	46.0	45.0
TVAL	754	99.0	100	99.0	99.0
Winter2	669	4.5	2.6	7.0	7.0

Table 4
Success rates (%) of function $S(\Phi)$ with all reflections up to $d_{\rm min} = 1$ Å.

Most symbols as in Table 2. $S_{\rm exp}$ is defined by $2(\langle|E|^2\rangle_{\mathbf{H}} - \langle|E|\rangle^2_{\mathbf{H}})$. [The theoretical $S_{\rm exp}$ for an acentric $|E|$ distribution is 0.43.] It can be seen how the success rates of $S(\Phi)$ are excellent.

Code	$\|E\|_{\rm min}$	Φ	Total no. reflections used	$S_{\rm exp}$	No. sets	% solutions	r.n.
PGE2	1.35	138	1098	0.532	100	100	1
MBH2	1.35	352	2312	0.486	100	100	1
SUOA	1.35	309	1862	0.520	500	3.2	2
Goldman2	1.35	248	1821	0.436	100	93	1
MGHEX	1.35	505	3712	0.414	100	41	3
Munich1	1.35	250	1611	0.528	500	6.4	4
BHAT	1.35	176	1147	0.446	100	50	2
TVAL	1.35	559	4972	0.368	100	98	1
Winter2	1.35	523	4120	0.468	500	4.0	1
APAPA	1.35	449	2731	0.574	500	2.0	7
	1.50	307	2731	0.574	500	3.8	6
PEP1	1.35	528	3829	0.482	500	4.8	25
	1.45	407	3829	0.482	500	6.0	3

value of $S(\Phi)$ is reached. The R values in Table 2 have been computed with

$$R = 1000\left\{1 - \left[\sum(F_oF_c)^{1/2}\right]^2 \Big/ \left(\sum F_o \sum F_c\right)\right\} \qquad (17)$$

assuming an overall thermal parameter. A solution has been considered as correct if after 10 Fourier cycles most atoms show up in the E map. The number of cycles was enlarged to 40 for the largest structures tested (and for PGE2).

The three analysed $|E|_{\rm min}$ values were 1.25, 1.45 and 1.65. Inspection of the data listed in Table 2 clearly shows that the best results are obtained for $|E|_{\rm min} = 1.25$. Not only the ranking number of the correct solutions is lower but also the number of correct solutions increases. In most cases, the correct solution is the top solution. This result is understandable since lower $|E|_{\rm min}$ values produce more accurate $|E(\mathbf{H}, \mathbf{\Phi})|$ and, owing to the size of these structures, the increase in the number of variables is still manageable.

For the larger structures, only tests with $|E|_{\rm min}$ in the range 1.45–1.59 have been performed owing to the large number of triplets. For APP, rubredoxin and alpha1, the results look very promising. One surprise has been the rather low ranking number of the correct solutions for these three compounds. For pheromone, however, the ranking number of the correct solution is 65 (out of 4000 trials), which could be identified because it has the best PSIZERO and RESID figures of merit. Tests with larger structures possessing the origin fixed in all three directions, *e.g.* $P2_12_12_1$, have been avoided. As shown in Rius *et al.* (1994), the number of correct solutions drastically decreases in comparison with space groups having the origin floating at least in one direction. Since the maximization procedure used by the actual version of *XLENS*™ is a conventional tangent formula refinement, it is very improbable that correct solutions can be found for these compounds.

4.2. Function $S(\Phi)$ for different reflection ratios $r = N_{\rm weak}/N_{\rm large}$

Since the first applications of the $S(\Phi)$ function, the number of weak reflections has been estimated by making $N_{\rm weak} \cong N_{\rm large}$. Owing to the good results obtained from the beginning, no further investigations were carried out regarding this point. Recently, in order to complete the study of the function, *XLENS*™ has been slightly modified to allow for different ratios r. The results of a series of test calculations for $|E|_{\rm min} = 1.25$ with ratios ranging from 0.5 to 2 are listed in Table 3. These results indicate that the efficiency of the function is rather insensitive to ratio variations within the studied interval.

Finally, a series of test calculations has been performed including not only the large and weak reflections *but all measured reflections*. As can be seen in Table 4, the results are also excellent, although the practical importance of these results is limited because the computing effort is much higher than for lower ratios.

Work supported by the Ministerio de Educacíon y Cultura (Project PB98-0483).

References

Anderson, D. H., Weiss, M. S. & Eisenberg, D. (1996). *Acta Cryst.* D**52**, 469–480.

Antel, J., Sheldrick, G. M., Bats, J. W., Kessler, H. & Müller, A. (1995). Unpublished.

Bhat, T. N. & Ammon, H. L. (1990). *Acta Cryst.* C**46**, 112–116.

Bhuiya, A. K. & Stanley, E. (1963). *Acta Cryst.* **16**, 981–984.

Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Spagna, R. (2000). *Acta Cryst.* A**56**, 451–457.

Butters, T., Hütter, P., Jung, G., Pauls, N., Schmitt, H., Sheldrick, G. M. & Winter, W. (1981). *Angew. Chem.* **93**, 904–905.

Debaerdemaeker, T., Tate, C. & Woolfson, M. M. (1985). *Acta Cryst.* A**41**, 286–290.

Debaerdemaeker, T. & Woolfson, M. M. (1983). *Acta Cryst.* A**39**, 193–196.

DeTitta, G. T., Langs, D. A., Edmonds, J. W. & Duax, W. L. (1980). *Acta Cryst.* B**36**, 638–645.

Glover, I., Haneef, I., Pitts, J.-E., Wood, S. P., Moss, D., Tickle, I. J. & Blundell, T. L. (1983). *Biopolymers*, **22**, 293–304.
Irngartinger, H., Reibel, W. R. K. & Sheldrick, G. M. (1981). *Acta Cryst.* B**37**, 1768–1771.
Karle, J. & Hauptman, H. (1956). *Acta Cryst.* **9**, 635–651.
Oliver, J. D. & Strickland, L. C. (1984). *Acta Cryst.* C**40**, 820–824.
Pedio, M., Felici, R., Torrelles, X., Rudolf, P., Capozi, M., Rius, J. & Ferrer, S. (2000). *Phys. Rev. Lett.* **85**, 1040–1043.
Poyser, J. P., Edwards, P. L., Anderson, J. R., Hursthouse, M. B., Walker, N. P., Sheldrick, G. M. & Whalley, A. J. S. (1986). *J. Antibiot.* **39**, 167–169.
Privé, G. G., Anderson, D. H., Wesson, L., Cascio, D. & Eisenberg, D. (1999). *Protein Sci.* **8**, 1–9.
Rius, J. (1993). *Acta Cryst.* A**49**, 406–409.
Rius, J. (2000). *Powder Diffr.* **14**, 267–273.
Rius, J. & Miravitlles, C. (1991). *Acta Cryst.* A**47**, 567–571.
Rius, J., Miravitlles, C. & Allmann, R. (1996). *Acta Cryst.* A**52**, 634–639.
Rius, J., Miravitlles, C., Gies, H. & Amigó, J. M. (1999). *J. Appl. Cryst.* **32**, 89–97.
Rius, J., Sañé, J., Miravitlles, C., Amigó, J. M., Reventós, M. M. (1994). *Acta Cryst.* A**51**, 268–270.
Rius, J., Sañé, J., Miravitlles, C., Gies, H., Marler, B. & Oberhagemann, U. (1995). *Acta Cryst.* A**51**, 840–845.
Rius, J., Torrelles, X. & Miravitlles, C. (2000). EPDIC-7: Abstracts Book, p. 59.
Rius, J., Torrelles, X., Miravitlles, C., Ochando, L. E., Reventós, M. M. & Amigó, J. M. (2000). *J. Appl. Cryst.* **33**, 1208–1211.
Sheldrick, G. M., Dauter, Z., Wilson, K. S., Hope, H. & Sieker, L. C. (1993). *Acta Cryst*. D**49**, 18–23.
Smith, G. D., Duax, W. L., Langs, D. A., DeTitta, G. T., Edmonds, J. W., Rohrer, D. C. & Weeks, J. (1975). *J. Am. Chem. Soc.* **97**, 7242–7247.
Suck, D., Manor, P. C. & Saenger, W. (1976). *Acta Cryst.* B**32**, 1727–1737.
Szeimies-Seebach, U., Harnisch, J., Szeimies, G., Van Meersche, M., Germain, G. & Declercq, J. P. (1978). *Angew. Chem. Int. Ed. Engl.* **17**, 848–850.
Torrelles, X., Rius, J., Boscherini, F., Heun, S., Mueller, B. H., Ferrer, S., Alvarez, J. & Miravitlles, C. (1998). *Phys. Rev. B*, **57**, R4281–4284.
Torrelles, X., Rius, J., Miravitlles, C. & Ferrer, S. (1998). *Surf. Sci.* **423**, 338–345.
Weeks, C. M., DeTitta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). *Acta Cryst.* A**50**, 210–220.
Weeks, C. M., DeTitta, G. T., Miller, R. & Hauptman, H. A. (1993). *Acta Cryst.* D**49**, 179–181.
Williams, D. J. & Lawton, D. (1975). *Tetrahedron Lett.* pp. 111–114.
Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
Woolfson, M. M. (1970). *An Introduction to X-ray Crystallography*, pp. 234–240. Cambridge University Press.
Yao, J. X. (1981). *Acta Cryst.* A**37**, 642–644.